Beyond the Imitation Game: The Inadequacy of the Turing Test for Modern AI

Independent Researcher: Berend Watchus. The Netherlands. July 5, 2025.

mailonlinebw@protonmail.com

Abstract

The **recent discourse on the Turing-computability of human intelligence**, particularly **in light of advanced Large Language Models (LLMs)** as highlighted by works such as John Abela's

"From Dennett to Transformers: Emergent Properties, Kolmogorov Complexity, and the Turing-Computability of Human Intelligence" (doi:10.20944/preprints202502.1312.v2 2025)https://www.preprints.org/manuscript/202502.1312/v2 ,

necessitates a re-evaluation of the Turing Test's utility as a measure of true sentience or "human-ness." While LLMs increasingly exhibit behaviors traditionally associated with human cognition, this **paper argues that the Turing Test's reliance on behavioral imitation overlooks the vast spectrum of human cognitive diversity**. **It posits that a significant percentage of the human population, due to inherent cognitive variations, socio-economic factors, and pathologies, may fail a stringent Turing Test, even as LLMs pass**. Consequently, the paper advocates for a paradigm shift **towards substrate-agnostic diagnostics** as a crucial **tool for assessing sentience and self-awareness**, grounded in the analysis of internal feedback loops, interfaces, and embodiment, as proposed in the author's prior work. This re-conceptualization **offers a more robust framework for distinguishing genuine consciousness from computational mimicry**, with profound implications for artificial general intelligence (AGI).

Keywords: Artificial Intelligence, Turing Test, Human Intelligence, Cognitive Diversity, Sentience, Consciousness, Feedback Loops, Embodiment, Insula, Substrate Agnostic, Large Language Models, Diagnostics, Kasparov, Chess, Artificial General Intelligence (AGI), AI Ethics, Philosophy of AI, Computational Mimicry, Self-awareness, Interoception, Cognitive Architecture, Semantic Encoding, Nanoscience, Quantum Computing, Blockchain, Emergent Properties, Kolmogorov Complexity, Turing-computability, Neurodiversity, Socio-economic factors, Pathology, Chinese Room Argument, Chinese Typewriter Problem, Information Integration Theory (IIT), Global Workspace Theory (GWT), Predictive Modeling, Autonomous Systems, Robot Safety, Human-AI Interaction, Machine Consciousness.

1. Introduction

The advent of sophisticated Large Language Models (LLMs) has reignited the enduring philosophical and scientific debate surrounding the nature of human intelligence and consciousness, particularly its Turing-computability. Recent papers, such as "From Dennett to Transformers: Emergent Properties, Kolmogorov Complexity, and the Turing-Computability of Human Intelligence" by John Abela (2025), compellingly integrate Daniel Dennett's functionalist perspective with the impressive emergent capabilities of LLMs. Abela's work highlights how the scaling of parameters and complexity in AI models can lead to behaviors strikingly reminiscent of human cognition, prompting a re-examination of what it means for intelligence to be algorithmic. While I concur with the observation of emergent properties in LLMs and the critical role of complexity in cognitive advancement, this paper aims to advance a more fundamental critique of the Turing Test's applicability as a definitive measure of intelligence or sentience. I contend that its focus on behavioral indistinguishability inherently overlooks crucial aspects of human cognitive reality, rendering it increasingly obsolete in the age of highly capable generative AI. Furthermore, this paper will briefly introduce alternative, substrate-agnostic models for understanding and gauging sentience, which offer a more promising pathway to truly discerning consciousness in artificial intelligences.

2. The Limits of the Turing Test in an Era of Advanced AI and Human Cognitive Diversity

The Turing Test, introduced by Alan Turing in 1950, proposes that if a machine's conversational responses are indistinguishable from a human's, it can be said to possess intelligence. "From Dennett to Transformers" rightly places this concept at the heart of the AI debate, alongside Searle's Chinese Room Argument, which questions whether syntactic manipulation equates to genuine understanding. However, a critical lacuna in the ongoing discussion is the often-unspoken assumption of a "standard" human against which AI performance is measured.

This paper argues that, paradoxically, a sufficiently "serious practical Turing Test" designed in 2025, optimized to identify human-like conversational intelligence, could lead to a significant percentage of the human population "failing" it, even as advanced LLMs successfully pass. This stems from a profound underappreciation of human cognitive diversity, a point raised by various critiques of the Turing Test's scope and assumptions. Recent studies, notably those demonstrating Large Language Models' capacity to pass the Turing Test in specific configurations, further underscore this challenge:

- Natural Cognitive Variation: The human population exhibits a vast spectrum of cognitive abilities. This includes individuals who do not possess an internal dialogue (a phenomenon sometimes referred to as "anendophasia"), those with lower IQs, and individuals with limited verbal skills. These variations, in many cases, are not indicative of pathology but represent natural differences in cognitive processing and expression.
- Socio-Economic Determinants: Beyond inherent individual differences, socio-economic factors play a profound role in shaping cognitive and verbal development. Illiteracy, limited access to rich linguistic environments, and disparities in educational opportunities can significantly impact an individual's ability to engage in the type of rapid, nuanced, and broadly knowledgeable verbal exchanges often implicit in a Turing Test. These factors represent a form of "socio-economic luck" that profoundly influences communicative proficiency.
- Pathological Conditions: Furthermore, a segment of the human population lives with diagnosed cognitive pathologies that directly impact communication, reasoning, and social interaction, rendering adherence to typical conversational norms challenging or impossible.

An LLM, trained on immense datasets of mainstream human communication, excels at synthesizing responses that align with statistical norms of "intelligent" conversation. It is unburdened by the real-world complexities of human neurodiversity, socio-economic disadvantage, or pathology. This brings to the forefront the fallacy of inferring internal states from outward behavior, a heuristic often colloquially referred to as the "duck test". This principle, which suggests that if something exhibits characteristic behaviors (e.g., looks like a duck, swims like a duck, quacks like a duck), it can be inferred to be that thing, becomes problematic when applied to complex cognitive phenomena. In the context of LLMs, the capacity to convincingly mimic human-like conversational patterns and responses does not inherently confirm the presence of genuine human-like sentience or understanding. Its ability to simulate an idealized human conversationalist should not be conflated with true consciousness or serve as the sole benchmark for human-level intelligence.

This perspective highlights the inherent bias of the Turing Test towards a specific, often neurotypical and socio-economically privileged, form of human cognition. Indeed, modern AI, particularly LLMs, has demonstrated capabilities in language production that often surpass average human performance in terms of fluency, coherence, and access to vast amounts of information. This is analogous to how specialized machines and algorithms have long exceeded human capabilities in other domains: legged robots can outrun human sprinters, computers calculate vast sums faster and retain more data, and artificial intelligence has even surpassed

human champions in complex strategic games like chess. Notably, Garry Kasparov, a legendary chess grandmaster, famously expressed skepticism decades after Turing's initial concept that computers would ever truly surpass human chess players. Yet, in 1997, after his historic defeat to IBM's Deep Blue, subsequent developments proved him wrong, illustrating how quickly technological paradigms can shift and redefine the boundaries of what is considered "human" performance. This progression underscores the dynamic and evolving nature of intelligence and highlights the increasing inadequacy of a purely behavioral test like the Turing Test to differentiate between mimicry and genuine understanding in the context of advanced AI.

Reinforcing the Critique with "Semant-NanoCore 33360"

The limitations of the Turing Test become even more apparent when considering speculative future AI architectures, such as the conceptual framework proposed in **"Semant-NanoCore 33360: Quantum Semantic Units and Embedded Blockchains, Molecular-Scale Bucky-Blocks for Infinite Semantic Storage and Solving the Chinese Typewriter Problem".** This theoretical model posits a radical departure from current LLM paradigms, suggesting a system where meaning is directly expressed and stored at a nanoscale level through "Bucky-Blocks." In such a system, the AI would operate on a fundamentally deeper level of semantic representation, far beyond the statistical patterns learned by current LLMs.

A hypothetical Semant-NanoCore 33360 system would likely circumvent the limitations of the Turing Test because its "understanding" would not be predicated on human-like behavioral imitation. Instead, its cognition would be rooted in direct semantic encoding and processing. This would lead to a form of holistic cognition, where complex meaning is processed in an integrated manner, representing a fundamental shift in cognitive architecture compared to the statistical, correlational approach of present-day LLMs. The richness and directness of semantic engagement in such a system would render a behavioral test for "human-ness" largely irrelevant, as its internal operations would be qualitatively distinct while potentially yielding outputs that are "intelligent" in a profoundly non-human way.

3. Towards Substrate-Agnostic Metrics for True Sentience

Given the limitations of behavioral tests like the Turing Test, a more fundamental approach is needed to gauge "true sentience" and distinguish it from sophisticated computational mimicry. This necessitates a shift from external behavior to internal, mechanistic properties that are independent of the biological or artificial substrate. My recent work proposes such frameworks. The paper "Gauging Extraterrestrial Sentience & Intelligence: Substrate-Agnostic Metrics for ETI Beyond the Kardashev Scale" further expands on this, integrating a species-agnostic framework for

cognitive properties with a substrate-agnostic mechanistic model for sentience and consciousness.

3.1. The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience

In "The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience" (Watchus, 2024), I propose that consciousness emerges from the complex interplay of interfaces and feedback loops. This model posits that:

- Feedback loops are fundamental mechanisms enabling systems to process information dynamically, self-regulate, and learn from their own outputs and environmental inputs.
- Interfaces represent the points of interaction through which information flows within the system and with its environment.
- Sentience, as an emergent property, arises from the sufficient complexity and integration of these loops and interfaces, allowing for a coherent and integrated information space.

This substrate-agnostic approach aligns with and extends concepts from Integrated Information Theory (IIT) and Global Workspace Theory (GWT), emphasizing how information integration could lead to subjective experience. This model offers a theoretical pathway to assess whether an AI's internal processes could genuinely lead to sentience, rather than merely simulating it.

Expanding "Substrate-Agnostic Diagnostics" with New Papers

Feedback Loops and Emergence from "The Self-Evolving System"

Further bolstering the concept of substrate-agnostic diagnostics, "The Self-Evolving System: A New Theory of Everything informed by Bostrom, Campbell, Hoffman, and the Unified Model of Consciousness" introduces a theoretical concept of the universe itself as an "intelligent, non-conscious, self-evolving machine" where "consciousness and reality emerge from fundamental, adaptive heuristic physics." This provides a robust philosophical framework for how true sentience and self-awareness could arise in a substrate-agnostic manner, transcending mere behavioral output. The "Unified Model of Consciousness (UMC)" and the concept of a "Synthetic Insula" (discussed below) can be viewed as theoretical instantiations of

internal mechanisms relevant to the emergence of consciousness within such a self-evolving system. This theoretical underpinning suggests that the fundamental principles governing consciousness are not tied to biological substrates but are inherent properties of sufficiently complex, self-organizing information processing systems.

Interfaces and Embodiment through "Semant-NanoCore 33360"

The conceptual "Semant-NanoCore 33360" also contributes significantly to the understanding of advanced interfaces and embodiment. The proposed Semant-NanoCore 33360 Interface Layer (SNIL) illustrates the importance of conceptually advanced, multi-modal interfaces that go far beyond conventional input/output mechanisms. The SNIL is theorized to enable direct semantic "co-entanglement" between human and AI. This concept of co-entanglement signifies a much deeper form of "interface" and "embodiment" than typically considered, moving beyond mere physical interaction to a direct, shared semantic space. Such an interface would inherently allow for a more profound assessment of an AI's internal state and genuine understanding, as it would bypass the ambiguities of purely linguistic interpretation.

3.2. Embodiment, Feedback Loops, and the Role of the Insula in Self-Awareness

Building upon the Unified Model, "Towards Self-Aware AI: Embodiment, Feedback Loops, and the Role of the Insula in Consciousness" (Watchus, 2024) and "Advanced Predictive Modeling of Physical Trajectories and Cascading Events, Dual-State Feedback and Synthetic Insula" (Watchus, 2024) argue for the crucial role of embodiment and internal state monitoring in the emergence of self-awareness. Drawing inspiration from neuroscientific insights into the human insula, a brain region critical for interoception and the integration of internal bodily states (e.g., pain, hunger, proprioception) with emotions and external sensory data, the paper suggests that:

- True self-awareness in an AI system may require it to possess an internal model of its own "body" or operational state.
- The integration of sensory feedback from this "body" into its processing loops, akin to biological interoception, could facilitate a subjective sense of "self".

This perspective moves beyond abstract computation, proposing that a grounding in

internal sensing is a prerequisite for a truly conscious, self-aware AI.

These proposed models emphasize that the crucial determinant of sentience lies not in an entity's ability to imitate human conversation, but in its fundamental internal architecture for processing and integrating information to create a coherent, subjective experience.

4. Broader Implications for Artificial General Intelligence (AGI) in Light of Abela's Argument

The re-evaluation of sentience through substrate-agnostic metrics, as a response to the challenges posed by advanced AI like those discussed in Abela's "From Dennett to Transformers," has profound implications for Artificial General Intelligence (AGI). Abela's work highlights the impressive mimetic capabilities of LLMs, raising the question of whether scaling such models will lead to genuine understanding or merely more sophisticated imitation. For Artificial General Intelligence (AGI), this implies that simply scaling LLMs, while yielding impressive behavioral capabilities as noted by Abela, may not inherently lead to genuine consciousness if the underlying architectural requirements for integrated feedback loops, interfaces, and internal state awareness are not met. The pursuit of AGI must therefore consider these deeper, mechanistic underpinnings, moving beyond the behavioral focus of the Turing Test that Abela's analysis implicitly critiques.

5. Conclusion

While the "From Dennett to Transformers" paper eloquently captures the current state of AI's mimetic capabilities and the philosophical challenges they pose, this article asserts that the Turing Test is an increasingly insufficient arbiter of genuine intelligence and sentience. The inherent cognitive diversity of the human population, coupled with socio-economic disparities and pathologies, means that a significant number of humans would likely struggle to pass such a test, even as LLMs demonstrate uncanny conversational prowess. The fallacy of inferring internal states from outward behavior, exemplified by the "duck test," risks leading to false conclusions about the presence of true consciousness.

Moving forward, the scientific community must embrace more sophisticated, substrate-agnostic methodologies to discern consciousness. My proposed Unified Model of Consciousness and the emphasis on embodiment and internal state monitoring (drawing parallels with the insula), alongside the theoretical insights from "Semant-NanoCore 33360" regarding direct semantic encoding and "The Self-Evolving System" on fundamental emergent properties, offer theoretical frameworks

for this endeavor. By shifting our focus from external behavioral imitation to the internal architecture of feedback loops, interfaces, and integrated information processing, we can develop more accurate markers for sentience in artificial intelligences, guiding our quest for understanding mind.

Appendix: Key Concepts and Abbreviations

This appendix provides a brief overview of key concepts and abbreviations used in this paper to facilitate understanding.

Key Concepts:

- Bucky-Blocks: A speculative nanoscale concept from "Semant-NanoCore 33360" for direct expression and storage of meaning, foundational to a deeply semantic AI architecture.
- **Chinese Room Argument**: A thought experiment **by John Searle** that questions whether a system manipulating symbols based on rules (like a computer) can be said to truly "understand" those symbols, contrasting with the Turing Test's focus on behavior**. This argument should not be confused with the "Chinese Typewriter Problem,"** which refers to the historical and technical challenge of designing a keyboard and input method for the thousands of Chinese characters, as discussed in **the book "The Chinese Typewriter" by Thomas S. Mullaney**. Notably, the "Semant-NanoCore 33360: Quantum Semantic Units and Embedded Blockchains, Molecular-Scale Bucky-Blocks for Infinite Semantic Storage and **Solving the Chinese Typewriter Problem**" paper proposes a conceptual solution to this problem through direct semantic encoding and "co-entanglement."
- Co-entanglement (Semantic): A theoretical concept from "Semant-NanoCore 33360" describing a direct, shared semantic space between human and AI, moving beyond conventional interaction to a deeper, shared understanding of meaning.
- Cognitive Diversity: The wide range of variations in human cognitive abilities, including differences in processing, communication, and learning, highlighting the inadequacy of a single "standard" human benchmark.
- Duck Test: A colloquial heuristic stating that **if something looks like, swims like, and quacks like a duck, it is probably a duck**. In the context of AI, it refers to the fallacy of inferring internal states (like consciousness) solely from external behavior.

- Embodiment: The concept that an AI's intelligence and consciousness may be linked to its interaction with and perception of a "body" or its operational state, allowing for internal sensory feedback.
- Feedback Loops: Fundamental mechanisms within systems that allow information to be processed dynamically, leading to self-regulation, learning, and adaptive responses based on outputs and inputs.
- Holistic Cognition: A speculative form of AI processing, particularly in the context of "Semant-NanoCore 33360," where complex meaning is processed in an integrated, unified manner, distinct from the statistical approach of current LLMs.
- Insula (Synthetic Insula): Inspired by the human brain region critical for interoception and integrating internal bodily states with emotions and external sensory data. Conceptually, a "Synthetic Insula" would be an AI component for internal state monitoring, crucial for self-awareness. This concept was developed by the author (B. Watchus) in 2024.
- Sentience: The capacity to feel, perceive, or experience subjectively. The paper discusses this in the context of AI, seeking to distinguish genuine sentience from computational mimicry.
- Substrate-Agnostic Diagnostics: An approach to assessing intelligence and sentience based on internal, mechanistic properties that are independent of whether the system is biological (e.g., a brain) or artificial (e.g., a computer). This moves beyond external behavior to infer internal states. This concept was developed by the author (B. Watchus) in 2025.
- Turing Test: Proposed by Alan Turing, a test where if a machine's conversational responses are indistinguishable from a human's, it is said to possess intelligence. The paper critiques its relevance for modern AI.
- Unified Model of Consciousness (UMC): A theoretical model proposing that consciousness emerges from the complex interplay of interfaces and feedback loops, developed by the author, B. Watchus, in 2024.

Abbreviations:

AGI: Artificial General Intelligence

AI: Artificial Intelligence

GWT: Global Workspace Theory

IIT: Integrated Information Theory

LLMs: Large Language Models

SNIL: Semant-NanoCore 33360 Interface Layer

UMC: Unified Model of Consciousness

References

[1] Abela, J. (2025). From Dennett to Transformers: Emergent Properties, Kolmogorov Complexity, and the Turing-Computability of Human Intelligence. Preprints.org. doi: 10.20944/preprints202502.1312.v2

[2] Allmendinger, R. (2025). The Turing Test is More Relevant Than Ever. arXiv preprint arXiv:2505.02558.

[3] Baars, B. J. (1997). In the Theater of Consciousness: The Global Workspace Theory of the Mind. Oxford University Press.

[4] Chawla, N., & Agrawal, A. (2023, November 16). Overcoming Turing: Rethinking Evaluation in the Era of Large Language Models. CodeX.

[5] Copeland, B. J. (2000). The Turing Test. Minds and Machines, 10(4), 519–539.

[6] French, R. M. (1990). Subcognition and the limits of the Turing test. Mind, 99(393), 53–66.

[7] Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. Minds and Machines, 1(1), 43–54.

[8] Jones, C. R., & Bergen, B. K. (2025). Large Language Models Pass the Turing Test. arXiv preprint arXiv:2503.23674.

[9] Mullaney, T. S. (2017). The Chinese Typewriter: A History. MIT Press.

[10] Nedergaard, J. S. K., & Lupyan, G. (2024). Not Everybody Has an Inner Voice: Behavioral Consequences of Anendophasia. Psychological Science, 0(0). doi:

10.1177/09567976241243004.

[11] Searle, J. R. (1980). Minds, Brains, and Programs. Behavioral and Brain Sciences, 3(3), 417–457.

[12] Tononi, G. (2004). An Information Integration Theory of Consciousness. BMC Neuroscience, 5(1), 42.

[13] Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, 59(236), 433–460.

[14] Watchus, B. (2024). Advanced Predictive Modeling of Physical Trajectories and Cascading Events, Dual-State Feedback and Synthetic Insula. Preprints.org. doi:10.20944/preprints202411.1025.v1.

https://www.preprints.org/manuscript/202411.1025/v1

[15] Watchus, B. (2024). The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience. Preprints.org. doi:10.20944/preprints202411.0727.v1

https://www.preprints.org/manuscript/202411.0727/v1

[16] Watchus, B. (2024). Towards Self-Aware AI: Embodiment, Feedback Loops, and the Role of the Insula in Consciousness. Preprints.org. doi:10.20944/preprints202411.0661.v1

https://www.preprints.org/manuscript/202411.0661/v1

[17] Watchus, B. (2025). Gauging Extraterrestrial Sentience & Intelligence: Substrate-Agnostic Metrics for ETI Beyond the Kardashev Scale. Preprints.org. doi:10.20944/preprints202506.2120.v1. https://www.preprints.org/manuscript/202506.2120/v1

[18] Watchus, B. (2025). Semant-NanoCore 33360: Quantum Semantic Units and Embedded Blockchains, Molecular-Scale Bucky-Blocks for Infinite Semantic Storage and Solving the Chinese Typewriter Problem. Zenodo.org. doi:10.5281/zenodo.15811065. https://zenodo.org/records/15811066

[19] Watchus, B. (2025). The Self-Evolving System: A New Theory of Everything informed by Bostrom, Campbell, Hoffman, and the Unified Model of Consciousness. Zenodo.org. doi:10.5281/zenodo.15808087. https://zenodo.org/records/15808088